

RECEIVED
JAN 0 2 2003
TECH CENTER 1600/2300



1600

ENTERED

#13

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/917,384A

DATE: 12/18/2002

TIME: 09:29:29

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw

3 <110> APPLICANT: ADNEY, WILLIAM S.
4 DING, SHI-YOU
5 MCCARTER, SUZANNE
6 HIMMEL, MICHAEL E.
7 DECKER, STEPHEN R.
8 VINZANT, TODD B.
10 <120> TITLE OF INVENTION: THERMAL TOLERANT EXOGLUCANASE FROM ACIDOTHERMUS
11 CELLULOLYTICUS
13 <130> FILE REFERENCE: NREL 01-38
15 <140> CURRENT APPLICATION NUMBER: 09/917,384A
16 <141> CURRENT FILING DATE: 2001-07-28
18 <160> NUMBER OF SEQ ID NOS: 11
20 <170> SOFTWARE: PatentIn Ver. 2.1
22 <210> SEQ ID NO: 1
23 <211> LENGTH: 1121
24 <212> TYPE: PRT
25 <213> ORGANISM: Acidothermus cellulolyticus
27 <400> SEQUENCE: 1
28 Met Pro Gly Leu Arg Arg Arg Leu Arg Ala Gly Ile Val Ser Ala Ala
29 1 5 10 15
31 Ala Leu Gly Ser Leu Val Ser Gly Leu Val Ala Val Ala Pro Val Ala
32 20 25 30
34 His Ala Ala Val Thr Leu Lys Ala Gln Tyr Lys Asn Asn Asp Ser Ala
35 35 40 45
37 Pro Ser Asp Asn Gln Ile Lys Pro Gly Leu Gln Leu Val Asn Thr Gly
38 50 55 60
40 Ser Ser Ser Val Asp Leu Ser Thr Val Thr Val Arg Tyr Trp Phe Thr
41 65 70 75 80
43 Arg Asp Gly Gly Ser Thr Leu Val Tyr Asn Cys Asp Trp Ala Ala
44 85 90 95
46 Met Gly Cys Gly Asn Ile Arg Ala Ser Phe Gly Ser Val Asn Pro Ala
47 100 105 110
49 Thr Pro Thr Ala Asp Thr Tyr Leu Gln Leu Ser Phe Thr Gly Gly Thr
50 115 120 125
52 Leu Ala Ala Gly Gly Ser Thr Gly Glu Ile Gln Asn Arg Val Asn Lys
53 130 135 140
55 Ser Asp Trp Ser Asn Phe Asp Glu Thr Asn Asp Tyr Ser Tyr Gly Thr
56 145 150 155 160
58 Asn Thr Thr Phe Gln Asp Trp Thr Lys Val Thr Val Tyr Val Asn Gly
59 165 170 175
61 Val Leu Val Trp Gly Thr Glu Pro Ser Gly Ala Thr Ala Ser Pro Ser
62 180 185 190
64 Ala Ser Ala Thr Pro Ser Pro Ser Ser Ser Pro Thr Thr Ser Pro Ser

RAW SEQUENCE LISTING

DATE: 12/18/2002

PATENT APPLICATION: US/09/917,384A

TIME: 09:29:29

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw

```

65          195          200          205
67 Ser Ser Pro Ser Pro Ser Ser Ser Pro Thr Pro Thr Pro Ser Ser Ser
68      210          215          220
70 Ser Pro Pro Pro Ser Ser Asn Asp Pro Tyr Ile Gln Arg Phe Leu Thr
71 225          230          235          240
73 Met Tyr Asn Lys Ile His Asp Pro Ala Asn Gly Tyr Phe Ser Pro Gln
74          245          250          255
76 Gly Ile Pro Tyr His Ser Val Glu Thr Leu Ile Val Glu Ala Pro Asp
77          260          265          270
79 Tyr Gly His Glu Thr Thr Ser Glu Ala Tyr Ser Phe Trp Leu Trp Leu
80          275          280          285
82 Glu Ala Thr Tyr Gly Ala Val Thr Gly Asn Trp Thr Pro Phe Asn Asn
83          290          295          300
85 Ala Trp Thr Thr Met Glu Thr Tyr Met Ile Pro Gln His Ala Asp Gln
86 305          310          315          320
88 Pro Asn Asn Ala Ser Tyr Asn Pro Asn Ser Pro Ala Ser Tyr Ala Pro
89          325          330          335
91 Glu Glu Pro Leu Pro Ser Met Tyr Pro Val Ala Ile Asp Ser Ser Val
92          340          345          350
94 Pro Val Gly His Asp Pro Leu Ala Ala Glu Leu Gln Ser Thr Tyr Gly
95          355          360          365
97 Thr Pro Asp Ile Tyr Gly Met His Trp Leu Ala Asp Val Asp Asn Ile
98          370          375          380
100 Tyr Gly Tyr Gly Asp Ser Pro Gly Gly Gly Cys Glu Leu Gly Pro Ser
101 385          390          395          400
103 Ala Lys Gly Val Ser Tyr Ile Asn Thr Phe Gln Arg Gly Ser Gln Glu
104          405          410          415
106 Ser Val Trp Glu Thr Val Thr Gln Pro Thr Cys Asp Asn Gly Lys Tyr
107          420          425          430
109 Gly Gly Ala His Gly Tyr Val Asp Leu Phe Ile Gln Gly Ser Thr Pro
110          435          440          445
112 Pro Gln Trp Lys Tyr Thr Asp Ala Pro Asp Ala Asp Ala Arg Ala Val
113          450          455          460
115 Gln Ala Ala Tyr Trp Ala Tyr Thr Trp Ala Ser Ala Gln Gly Lys Ala
116 465          470          475          480
118 Ser Ala Ile Ala Pro Thr Ile Ala Lys Ala Ser Gln Thr Gly Asp Tyr
119          485          490          495
121 Leu Arg Tyr Ser Leu Phe Asp Lys Tyr Phe Lys Gln Val Gly Asn Cys
122          500          505          510
124 Tyr Pro Ala Ser Ser Cys Pro Gly Ala Thr Gly Arg Gln Ser Glu Thr
125          515          520          525
127 Tyr Leu Ile Gly Trp Tyr Tyr Ala Trp Gly Gly Ser Ser Gln Gly Trp
128          530          535          540
130 Ala Trp Arg Ile Gly Asp Gly Ala Ala His Phe Gly Tyr Gln Asn Pro
131 545          550          555          560
133 Leu Ala Ala Trp Ala Met Ser Asn Val Thr Pro Leu Ile Pro Leu Ser
134          565          570          575
136 Pro Thr Ala Lys Ser Asp Trp Ala Ala Ser Leu Gln Arg Gln Leu Glu
137          580          585          590

```

RAW SEQUENCE LISTING

DATE: 12/18/2002

PATENT APPLICATION: US/09/917,384A

TIME: 09:29:29

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw

```

139 Phe Tyr Gln Trp Leu Gln Ser Ala Glu Gly Ala Ile Ala Gly Gly Ala
140      595      600      605
142 Thr Asn Ser Trp Asn Gly Asn Tyr Gly Thr Pro Pro Ala Gly Asp Ser
143      610      615      620
145 Thr Phe Tyr Gly Met Ala Tyr Asp Trp Glu Pro Val Tyr His Asp Pro
146 625      630      635      640
148 Pro Ser Asn Asn Trp Phe Gly Phe Gln Ala Trp Ser Met Glu Arg Val
149      645      650      655
151 Ala Glu Tyr Tyr Tyr Val Thr Gly Asp Pro Lys Ala Lys Ala Leu Leu
152      660      665      670
154 Asp Lys Trp Val Ala Trp Val Lys Pro Asn Val Thr Thr Gly Ala Ser
155      675      680      685
157 Trp Ser Ile Pro Ser Asn Leu Ser Trp Ser Gly Gln Pro Asp Thr Trp
158      690      695      700
160 Asn Pro Ser Asn Pro Gly Thr Asn Ala Asn Leu His Val Thr Ile Thr
161 705      710      715      720
163 Ser Ser Gly Gln Asp Val Gly Val Ala Ala Ala Leu Ala Lys Thr Leu
164      725      730      735
166 Glu Tyr Tyr Ala Ala Lys Ser Gly Asp Thr Ala Ser Arg Asp Leu Ala
167      740      745      750
169 Lys Gly Leu Leu Asp Ser Met Trp Asn Asn Asp Gln Asp Ser Leu Gly
170      755      760      765
172 Val Ser Thr Pro Glu Thr Arg Thr Asp Tyr Ser Arg Phe Thr Gln Val
173      770      775      780
175 Tyr Asp Pro Thr Thr Gly Asp Gly Leu Tyr Ile Pro Ser Gly Trp Thr
176 785      790      795      800
178 Gly Thr Met Pro Asn Gly Asp Gln Ile Lys Pro Gly Ala Thr Phe Leu
179      805      810      815
181 Ser Ile Arg Ser Trp Tyr Thr Lys Asp Pro Gln Trp Ser Lys Val Gln
182      820      825      830
184 Ala Tyr Leu Asn Gly Gly Pro Ala Pro Thr Phe Asn Tyr His Arg Phe
185      835      840      845
187 Trp Ala Glu Ser Asp Phe Ala Met Ala Asn Ala Asp Phe Gly Met Leu
188      850      855      860
190 Phe Pro Ser Gly Ser Pro Ser Pro Thr Pro Ser Pro Thr Pro Thr Ser
191 865      870      875      880
193 Ser Pro Ser Pro Thr Pro Ser Ser Ser Pro Thr Pro Ser Pro Ser Pro
194      885      890      895
196 Ser Pro Thr Gly Asp Thr Thr Pro Pro Ser Val Pro Thr Gly Leu Gln
197      900      905      910
199 Val Thr Gly Thr Thr Thr Ser Ser Val Ser Leu Ser Trp Thr Ala Ser
200      915      920      925
202 Thr Asp Asn Val Gly Val Ala His Tyr Asn Val Tyr Arg Asn Gly Thr
203      930      935      940
205 Leu Val Gly Gln Pro Thr Ala Thr Ser Phe Thr Asp Thr Gly Leu Ala
206 945      950      955      960
208 Ala Gly Thr Ser Tyr Thr Tyr Thr Val Ala Ala Val Asp Ala Ala Gly
209      965      970      975
211 Asn Thr Ser Ala Gln Ser Phe Ala Gly Asp Ser Asp Asp Gly Ile Ala

```

RAW SEQUENCE LISTING

DATE: 12/18/2002

PATENT APPLICATION: US/09/917,384A

TIME: 09:29:29

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw

```

212          980          985          990
214 Val Ala Ser Pro Ser Pro Ser Pro Thr Pro Thr Ser Ser Pro Ser Pro
215          995          1000          1005
217 Thr Pro Ser Pro Thr Pro Ser Pro Thr Ser Thr Ser Gly Ala Ser Cys
218      1010          1015          1020
220 Thr Ala Thr Tyr Val Val Asn Ser Asp Trp Gly Ser Gly Phe Thr Thr
221 1025          1030          1035          1040
223 Thr Val Thr Val Thr Asn Thr Gly Thr Arg Ala Thr Ser Gly Trp Thr
224          1045          1050          1055
226 Val Thr Trp Ser Phe Ala Gly Asn Gln Thr Val Thr Asn Tyr Trp Asn
227          1060          1065          1070
229 Thr Ala Leu Thr Gln Ser Gly Lys Ser Val Thr Ala Lys Asn Leu Ser
230          1075          1080          1085
232 Tyr Asn Asn Val Ile Gln Pro Gly Gln Ser Thr Thr Phe Gly Phe Asn
233      1090          1095          1100
235 Gly Ser Tyr Ser Gly Thr Asn Thr Ala Pro Thr Leu Ser Cys Thr Ala
236 1105          1110          1115          1120
238 Ser
241 <210> SEQ ID NO: 2
242 <211> LENGTH: 3365
243 <212> TYPE: DNA
244 <213> ORGANISM: Acidothermus cellulolyticus
246 <400> SEQUENCE: 2
247 atgccaggat tacgacggcg actccgcgcc ggtatcgtct cggcggcgcc gttggggtcg 60
248 ctggttagcg ggctcggttc cgtcgcacca gtcgcgcacg cggcggtgac tctcaaagcg 120
249 cagtataaga acaatgattc ggcgccgagt gacaaccaga tcaaaccggg tctccagttg 180
250 gtgaataacc ggctcgctgc ggtggatttg tcgacggtga cggtgcggtg ctggttcacc 240
251 cgggatggtg ggctcgtcgac actggtgtac aactgtgact gggcggcgat ggggtgtggg 300
252 aatatccgcg cctcgttcgg ctcggtgaac cggcgacgc cgacggcgga cacctacctg 360
253 cagttgtcgt tcaactggtg aacgttggcc gctggtgggt cgacgggtga gattcaaac 420
254 cgggtgaata agagtgactg gtcgaacttt gatgagacca atgactactc gtatgggacg 480
255 aacaccacct tccaggactg gacgaaggtg acggtgtacg tcaacggcgt gttggtctgg 540
256 gggaccgaac cgtccggagc gacggcgtct ccatccgcgt cggcgacgcc cagcccgtcc 600
257 agttcaccga ccacgagtc gagttcgctc ccgtcgccga gcagcagccc gacgccgaca 660
258 ccgagcagct cgtcgccgcc ccgtcgctcca acgaccgta catccagcgg ttccctacga 720
259 tgtacaacaa gattcacgac ccagcgaacg gctacttcag cccgcaggga attccctacc 780
260 actcggtaga aacgctcatc gttgaggcac cggactacgg gcacgagaca acttcggagg 840
261 cgtacagctt ctggctctgg ctggaagcga cgtacggcgc agtgaccggc aactggacgc 900
262 cgttcaacaa cgcctggacg acgatggaaa cgtacatgat cccgcagcac gcggaccagc 960
263 cgaacaacgc gtcgtacaac cccaacagcc cggcgtcgta cgctccggaa gagccgctgc 1020
264 ccagcatgta cccggttgcc atcgacagca gcgtgccggt tgggcacgac ccgctcgccg 1080
265 ccgaattgca gtcgacgtac ggcactccgg acatttacgg catgcaactg ctggccgacg 1140
266 ttgacaacat ctacggatac ggcgacagcc cggcggtgg ttgcgaactc ggtccttcgg 1200
267 ctaaggcgct ctctacatc aacacattcc agcgggctc gcaggagtcc gtctgggaga 1260
268 cggtcaccca gccgacgtgc gacaacggca agtacggtgg ggcgcacggc tacgtcgacc 1320
269 tgttcatcca gggttcgacg ccgccgcagt ggaagtacac cgatgccccg gacgccgacg 1380
270 cccgtgccgt ccaggctgcg tactgggcct acacctgggc atcggcgag ggcaaggcaa 1440
271 gcgcgattgc cccgacgatc gccaaaggcg gccaaaccgg cgactacctg cggctactcg 1500
272 tctttgacaa gtacttcaag caggtcggca actgctaccc ggccagctcc tgccctggag 1560

```

RAW SEQUENCE LISTING

DATE: 12/18/2002

PATENT APPLICATION: US/09/917,384A

TIME: 09:29:29

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw

```

273 caaccggacg ccagagcgag acctacctga tcggctggta ctacgcctgg ggcggtcaa 1620
274 gccaaggctg ggccctggcg attggtgacg gcgccgcgca cttcggtac cagaatccgc 1680
275 ttgccgcgtg ggcgatgtcg aacgtgacac cgctcattcc gctctcgccc acggcaaaga 1740
276 gcgactgggc ggcgagcttg cagcgccagc tggagttcta ccagtggttg caatccgcgg 1800
277 aaggagccat tgcgggcggc gccaccaaca gctggaacgg caattacggg accccgcccg 1860
278 ccggagactc gaccttctac ggcattggcg acgactggga gccgggtctac cagcaccgcg 1920
279 cgagcaacaa ctggttcggc ttccaggcgt ggtccatgga acgggttgcc gagtactact 1980
280 acgtcaccgg cgaccgaag gccaggcgcg tgctcgacaa gtgggtcgca tgggtgaagc 2040
281 cgaatgtcac caccggtgcc tcatggtcga ttccgtcgaa tttgtcctgg agcggccaac 2100
282 cggatacctg gaatccgagc aaccaggaa cgaatgccaa cctgcacgtg accatcacgt 2160
283 cgtccgggca ggacgtcggt gttgccgcgg cgctcgcgaa gacactcgag tactacgcgg 2220
284 caaaatccgg cgatacggcc tcgcgcgacc tcgcgaaggg attgctcgac tccatgtgga 2280
285 acaacgacca ggacagcctc ggtgtgagca caccggagac gcggaccgac tactctcggt 2340
286 tcaactcagg gtacgacccg acgactggtg acggcctcta catcccgtcg ggttgacgg 2400
287 ggaccatgcc caacggtgac caaatcaagc cgggtgcgac cttcctgagc atccggctct 2460
288 ggtacaccaa ggatccgcag tggtcgaagg tgcaggcgta cctcaacggc gggcctgctc 2520
289 cgacgttcaa ctaccaccgg ttctgggagg agtccgactt cgcgatggcg aacgccgatt 2580
290 ttggcatgct cttcccatcc gggtcgccca gcccgacccc gagcccgact ccgacgtcgt 2640
291 ccccgagccc gactccgagc agctcgccga cgccgtcgcc cagcccgtca ccgaccggcg 2700
292 acaccacgcc gccgagcgtg ccgacgggtc ttcaggtcac cgggacaacg acgtcgtccg 2760
293 tgtcgctcag ctggaccgcg tccaccgaca acgtcggcgt cgcgactac aacgtgtacc 2820
294 gaaacggcac gctggtgggt cagccgacag cgacgtcggt caccgacacc ggccgtggctg 2880
295 ctggcacgtc gtacacgtac acagtggcgg ccgttgatgc ggccggtaac acgtcggcgc 2940
296 agagcttcgc cggtagacgc gacgacggca tcgcgtcgcg gagcccgctg ccgagcccga 3000
297 ctccgacgtc gtccccgagc ccaacgccgt cgccgacacc gtcaccgacg tccaccagcg 3060
298 gcgcacgtg cactgtctac tacgttgtca atagcgactg gggtagcggc ttcacgacaa 3120
299 ccgtgaccgt gacgaacacc ggcaccaggg ccaccagtgg ctggacggtc acgtggagct 3180
300 ttgccggtaa tcagacggtc accaactact ggaacaccgc gctgacgcaa tccggaaagt 3240
301 ccgtgaccgc aaagaacctg agttacaaca acgtcatcca acctggtcag tcgacgacct 3300
302 ttggattcaa cggaagttac tcaggaacaa acaccgcgcc gacgctcagc tgcacggcaa 3360
303 gctga 3365

```

306 <210> SEQ ID NO: 3

307 <211> LENGTH: 34

308 <212> TYPE: PRT

309 <213> ORGANISM: Acidothermus cellulolyticus

311 <400> SEQUENCE: 3

312 Met Pro Gly Leu Arg Arg Arg Leu Arg Ala Gly Ile Val Ser Ala Ala

313 1 5 10 15

315 Ala Leu Gly Ser Leu Val Ser Gly Leu Val Ala Val Ala Pro Val Ala

316 20 25 30

318 His Ala

321 <210> SEQ ID NO: 4

322 <211> LENGTH: 153

323 <212> TYPE: PRT

324 <213> ORGANISM: Acidothermus cellulolyticus

326 <400> SEQUENCE: 4

327 Ala Val Thr Leu Lys Ala Gln Tyr Lys Asn Asn Asp Ser Ala Pro Ser

328 1 5 10 15

330 Asp Asn Gln Ile Lys Pro Gly Leu Gln Leu Val Asn Thr Gly Ser Ser

VERIFICATION SUMMARY

DATE: 12/18/2002

PATENT APPLICATION: US/09/917,384A

TIME: 09:29:30

Input Set : A:\Nrel0138.app

Output Set: N:\CRF4\12182002\I917384A.raw